QUAERO/Experiment Internal Interface

Bruce Knuteson*
Massachusetts Institute of Technology
(Dated: May 7, 2005)

QUAERO has a streamlined internal interface to each frontier energy collider experiment. An experiment wishing to make its data available through QUAERO should provide four large text files. The first of these files (`data.txt`) contains the events observed in the experiment, reduced to the momenta of energetic final state objects ($e^{\pm}$, $\mu^{\pm}$, $\tau^{\pm}$, γ, b, and j); the second (`sm.txt`) contains weighted events corresponding to all Standard Model processes, as they would be observed in the detector; the third and fourth (`partonEvents.txt` and `recoEvents.txt`) contain a large sample of events that have been run through the detector simulation, written out at parton level and at the level of reconstructed objects, repectively. Specification of sources of experimental systematic uncertainties and their effects, and specification of any post-processing steps, completes the interface. Successful interfacing has been achieved so far with H1 in HERA I, with DØ in Tevatron Run I, and with ALEPH and L3 in LEP 2.

Contents

I. INTRODUCTION

It is generally recognized that the Standard Model, a successful description of the fundamental particles and their interactions, must be incomplete. Models that extend the Standard Model often predict rich phenomenology at the scale of a few hundred GeV, an energy regime accessible to the HERA, LEP, Tevatron, and LHC colliders. Due in part to the complexity of the apparatus required to test models at such large energies, experimental responses to these ideas have not kept pace. Any technique that reduces the time required to test a particular candidate theory would allow more such theories to be tested, reducing the possibility that the data contain overlooked evidence for new physics.

Once data are collected and the backgrounds have been understood, the testing of any specific model in principle follows a well-defined procedure. In practice, this process has been far from automatic. Even when the basic selection criteria and background estimates are taken from a previous analysis, the reinterpretation of the data in the context of a new model often requires a substantial length of time.

Ideally, the data should be "published" in such a way that others in the community can easily use those data to test a variety of models. The publishing of experimental distributions in journals allows this to occur at some level, but an effective publishing of a multidimensional data set by a large particle physics experiment has proven difficult. The problem appears to be that such data are context-specific, requiring detailed knowledge of the complexities of the apparatus. This knowledge must somehow be incorporated either into the data or into whatever tool the non-expert would use to analyze those data.

QUAERO is designed to enable the analysis of high energy collider data by non-experts. The original version of QUAERO [1], developed by the DØ experiment at Fermilab, computes cross section $\times$ branching ratio limits on new phenomena. QUAERO has recently been significantly extended, allowing simultaneous analysis of data from several frontier energy collider experiments. As the project has grown, a standard internal interface between QUAERO and each experiment has been defined. This article describes this interface.

In the interest of ease of manipulation, portability, and long-term maintenance, data are stored in ASCII text files. Events are reduced to 4-vectors of final state objects, the level at which it is customary to describe analyses in articles submitted to Physical Review Letters. Each row of a QUAERO text file represents one event, and consists of several strings and numbers separated by spaces or tabs.

Each object in each event is specified by the type of object and its momentum 4-vector. The object type is **e-** or **e+** for an electron or positron, **mu-** or **mu+** for a muon, **tau-** or **tau+** for a tau, **ph** for a photon, **j** for a non-b-tagged jet, **b** for a b-tagged jet, and **uncl** for unclustered energy. The momentum 4-vector of the object consists of its energy, polar angle, and azimuthal angle. All identified objects are energetic and isolated.

*Electronic address: knuteson@mit.edu; URL: `http://mit.fnal.gov/~knuteson/`

Collider	Object type	4-momentum format
LEP 2	j,b, `uncl`	m E $\cos\theta$ ϕ [rad]
LEP 2	e,μ,τ,γ	E $\cos\theta$ ϕ [rad]
Other	j,b, `uncl`	m p_T η ϕ [deg]
Other	e,μ,τ,γ	p_T η ϕ [deg]

TABLE I: Format for the specification of 4-momenta for objects at LEP 2 and other colliders. The mass m, energy E, and transverse momentum p_T of an object is specified in units of GeV. The mass m is only provided for jets, b-tagged jets, and unclustered energy, since the mass of the object is known for e, μ, τ, and γ. The polar direction is specified as the cosine of the polar angle θ in the case of lepton-lepton collisions, and in terms of the pseudorapidity η in lepton-hadron and in hadron-hadron collisions. The azimuthal angle ϕ is written in units of radians in lepton-lepton collisions, and in units of degrees in lepton-hadron and hadron-hadron collisions.

Missing energy ($\vec{\not p}$) is determined by energy conservation.

Associated with each event is a weight, chosen such that if all events are considered together, each with its appropriate weight, the correct distribution is obtained, with normalization equal to the total number of predicted events (in the case of signal or background) or observed events (in the case of data). Each event may have a different weight, allowing the use of Monte Carlos that produce weighted events. Each row in a QUAERO file begins with a string describing the type of event, followed by the run number and event number, the weight of the event, the type of collision, the center-of-mass energy of the collision, and the objects in the event. The end of an event is signaled by the presence of a semicolon, separated by whitespace.

Section II describes the components of the interface in detail sufficient for an experiment interested in participating in the QUAERO project to begin assembling the necessary ingredients. Section III summarizes.

II. INTERFACE

The components of the QUAERO/Experiment internal interface consists of five parts: data, background, detector simulation, systematic errors, and refinements.

1. Data

Observed events are provided in one large text file (`data.txt`), with event per one line, in the format described above. The 4-momentum of each object is specified in one of several forms, as shown in Table I, depending on the type of collision and the type of object. Each event also contains a few words of header information.

An event in a QUAERO file at LEP 2 containing two electrons and two b-tagged jets would look like this:

`eventType` `runNumber.eventNumber`
`weight` `collisionType` `sqrt(s)`
e+ E $\cos(\theta)$ ϕ **e-** E $\cos(\theta)$ ϕ
b m E $\cos(\theta)$ ϕ **b** m E $\cos(\theta)$ ϕ
uncl m E $\cos(\theta)$ ϕ **;**

where `collisionType` would be **e+e-**, `sqrt(s)` would be one of the nominal LEP 2 center of mass energies of 183, 189, 192, 196, 200, 202, 205, or 207, and ϕ is specified in units of radians. The event has been split across several lines due to the margin restrictions of the page; in the text file, the event is contained on a single line.

An event in a QUAERO file containing two electrons and two b-tagged jets at HERA, the Tevatron, or the LHC would look like this:

`eventType` `runNumber.eventNumber`
`weight` `collisionType` `sqrt(s)` `zvtx`
e+ p_T η ϕ **e-** p_T η ϕ
b m p_T η ϕ **b** m p_T η ϕ
uncl m p_T η ϕ **;**

where `collisionType` would be `e+p`, `e-p`, `ppbar`, or `pp`; `sqrt(s)` would be 301 or 319 for HERA I, 1800 for Tevatron Run I, 1960 for Tevatron Run II, and 14000 for the LHC; ϕ is specified in units of degrees; and `zvtx` is specified in units of centimeters. The collision point along the z axis must be specified in lepton-hadron and in hadron-hadron collisions.

In a QUAERO data file, `eventType` is replaced by the keyword `data`, and `weight` is set equal to 1. With actual numbers, an event in a data file from LEP 2 might look like this:

`data`	151942.19294	1	**e+e-**	189	
e+		45.2	+0.11	0.21	
e-		47.3	-0.05	3.56	
b	4.2	46.0	-0.16	1.71	
b	4.3	48.2	-0.02	4.90	
uncl	0.44	3.3	+0.07	3.97	**;**

2. Backgrounds

Weighted events corresponding to the Standard Model prediction are provided in a large file (`sm.txt`) with similar format. In `sm.txt`, the string `eventType` should be a string labeling the Standard Model process, and `weight` should be a number $\lesssim 0.01$ such that the sum of all weights in the file equals the total number of events predicted from the Standard Model.

3. Detector simulation

TURBOSIM [2] is used to create a fast detector simulation from the experiment's official detailed detector simulation. TurboSim uses events that have been run through the detailed simulation to create a large lookup table, which is then used to simulate the detector response to signal events. The construction of this lookup table requires two files, the first (`partonEvents.txt`) containing a sample of events at generator level, and the second (`recoEvents.txt`) containing the same events at the level of reconstructed objects. The weight of each event in `partonEvents.txt` and `recoEvents.txt` should be set equal to unity, but otherwise the format of these files is the same as that of `sm.txt`.

4. Systematic Errors

There are three aspects to the specification of systematic errors. The first of these is identification of the sources of systematic errors, and the assignment of a systematic error code to each source. New sources of systematic error are specified as additions to `systematicSources.txt` [3], which can be emailed to the Quaero authors. Sources of systematic uncertainty come in two different types: (m) measurement errors, instantiated as a number distributed as a Gaussian with mean 0 and width 1, and (i) identification errors, instantiated as a number distributed uniformly in the unit interval.

The second aspect to the specification of systematic errors is the assignment of the correlation of each new source of systematic error to the sources of systematic error that have already been specified. These correlations are specified as additions to `systematicCorrelations.txt` [4], which can be emailed to the Quaero authors.

The third aspect to the specification of systematic errors is the assignment of the effect each systematic error source has on the modeling of each event, including both the event's weight and the identification and 4-momentum determination of each object in the event. This should be sent in an email in English or in pseudo-code to the Quaero authors.

Uncertainties due to Monte Carlo statistics are incorporated by Quaero automatically.

5. Refinements

Post-processing steps should be described in an email in English or in pseudo-code. These steps will be converted into real code [5] by the Quaero authors. These post-processing steps may include identifying and removing objects that are outside fiducial volumes, removing events in final states contaminated by instrumental or environmental backgrounds not adequately modelled in `sm.txt`, and imposing the rule by which events are to be partitioned into final states.

III. SUMMARY

The interface between Quaero and a frontier energy collider experiment is at this point well specified. Each experiment provides its data in the form of the events collected in the detector, reduced to 4-momenta of energetic reconstructed objects; a sample of weighted events representing all Standard Model physics processes, similarly reduced to 4-momenta of energetic reconstructed objects; a sample of events that have been run through the detector simulation, both at generator level and at the level of reconstructed objects, which is used by TurboSim to construct a fast version of the detector simulation; specification of experimental systematic uncertainties and their effect; and specification of any post-processing that needs to be performed to ensure correct results.

The understanding represented by these items can then be used by Quaero to test specific hypotheses of how Nature behaves at the electroweak scale.

[1] DØ Collaboration (2001), *Phys. Rev. Lett.* **87** 231801.

[2] B. Knuteson (2004), TurboSim: A Self-Tuning Fast Detector Simulation; http://mit.fnal.gov/~knuteson/papers/turboSim.ps.

[3] B. Knuteson (2005), http://mit.fnal.gov/~knuteson/Quaero/quaero/doc/devel/systematicSources.txt.

[4] B. Knuteson (2005), http://mit.fnal.gov/~knuteson/Quaero/quaero/doc/devel/systematicCorrelations.txt.

[5] B. Knuteson (2005), http://mit.fnal.gov/~knuteson/Quaero/quaero/doc/devel/Refine.cc.